DOCUMENT RESUME

ED 300 452 TM 012 501

AUTHOR Kingston, Neal M.; And Others
TITLE The Graduate Record Examinations Validity Study Service: Yesterday and Tomorrow.
PUB DATE 86
NOTE 22p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Admission Criteria; Bayesian Statistics; *College Entrance Examinations; Departments; Graduate School Faculty; Higher Education; Least Squares Statistics; *Research Reports; Scores; *Selective Admission; Statistical Analysis; *Test Validity; *Validity
IDENTIFIERS Educational Testing Service; *GRE Validity Study Service

ABSTRACT

The Graduate Record Examinations (GRE) Validity Study Service (VSS) was developed by the Educational Testing Service at the request of the GRE Board to help graduate departments investigate the validity of their student selection procedures. The GRE VSS reports are aimed primarily at graduate department chairpersons and faculty members--those who select students for admission to graduate programs. A brief overview of the problems faced by graduate departments in selecting students is provided. Topics discussed include the following: (1) entities that use the GRE scores and validity information and purposes for which the scores and information are used; (2) types and forms of statistical information reported by the VSS; (3) VSS estimation methods; (4) statistical problems presented by subgroups of students, small samples, highly intercorrelated predictors, and restriction of range; and (5) types and uses of summary information (across institutions) provided. These topics are discussed in terms of past practices and future plans of the VSS. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED300452

THE GRADUATE RECORD EXAMINATIONS VALIDITY STUDY SERVICE: YESTERDAY AND TOMORROW*

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
NEAL M. KINGSTON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Neal M. Kingston
Robin V. Durso
Lisa M. Schneider

Educational Testing Service

*The ideas expressed in this paper are solely those of the authors and have not been endorsed by either the Graduate Record Examinations Board nor the Educational Testing Service.

TM012501

INTRODUCTION

The Graduate Record Examinations (GRE) Validity Study Service (VSS) was developed by Educational Testing Service at the request of the Graduate Record Examinations Board to help graduate departments investigate the validity of their graduate student selection procedures. The GRE VSS began operating in 1979 and has produced about 650 validity studies and three summary reports.

The GRE VSS operates in a particularly difficult environment. Institutional users vary considerably in their needs. Some departments admit almost all applicants, others select about one percent. Available criteria, most notably graduate first-year average, shows little variability within departments. An almost overwhelming problem has been that most departments admit only five to ten students a year. Even with combining data from several years, bivariate regression coefficients and zero-order correlations have huge sampling errors and multiple regression with three or more predictors is out of the question.

The GRE VSS has heretofore faced these problems by using a cooperative validity studies approach (Wilson, 1979). Individual studies were filled with caveats, but a summary report presented results for a large number of similar departments, such as humanities departments. To the extent that validity results were generalizable, and results indicated that they typically were, the average results provided general guidelines regarding the relative utility of various predictors.

The GRE VSS is now in the process of implementing empirical Bayes regression as the basis for future validity studies. Empirical Bayes regression (Braun & Jones, 1985) makes use of departmental similarities to provide accurate estimation of departmental regression weights, even when the sample sizes of individual departments are very small.

In this paper, we will answer the questions posed by the symposium organizer twice. Once for yesterday's GRE Validity Study Service, and once for the empirical Bayes based Validity Study Service of tomorrow.

THE GRE VSS YESTERDAY

WHO USES THE TEST SCORES, FOR WHAT PURPOSE? WHO WILL USE THE VALIDITY INFORMATION, FOR WHAT PURPOSE?

The following is excerpted from the Guide to the Use of the GRE Program (Educational Testing Service, 1986).

> The Graduate Record Examinations are designed to assess academic knowledge and skills relevant to graduate study, and the scores from the tests, when used properly, can improve the graduate education process.
>
> The GRE General Test is designed to measure broadly defined verbal, quantitative, and analytical skills. The Subject Tests have been developed to reflect material usually associated with undergraduate curricula, and they are particularly appropriate for educational uses in which the undergraduate subject area is relevant. Both the General Test and Subject Tests are appropriate for use in admitting students for graduate study and in making fellowship awards. They may also be used for predicting success in graduate school and for guidance in counseling students in their courses of graduate study.
>
> Other uses of the General Test or Subject Tests are inappropriate unless validated for that purpose by the user.
>
> Institutions using GRE scores in the admissions process are encouraged to examine the relationship between test scores and measures of performance in their academic programs. To assist in this, a Validity Study Service is available to user institutions free of charge through the GRE Program.

The actual users of GRE validity information varies considerably, but the GRE Validity Study Service (VSS) reports are aimed primarily at graduate department chairpersons and faculty members, that is, at those people who select students for admission to graduate programs. Other typical readers include staff in central admission offices and graduate deans. It is assumed that the typical GRE VSS participant has no background in statistics nor tests and measurements.

WHAT TYPES OF STATISTICAL INFORMATION DO YOU REPORT? IN WHAT FORM DO YOU REPORT THEM?

The chief form of statistical information reported in the current GRE VSS departmental report is the raw correlation between each predictor and each criterion. Since these correlations are typically based on samples of 25 to 40 students, they are supplemented by distributions of correlations from other similar departments that have participated in the GRE Validity Study Service (that is, biological sciences, humanities, physical and mathematical sciences, or social sciences). To the extent that validity of GRE scores and undergraduate grade point average is generalizable, the median of each distribution provides a good estimate of the predictive validity of the score in question. Departments are advised as to how similar they are to the other departments with which they have been grouped, so they can judge how likely it is that the central tendency of the distribution is representative of their department.

In addition, the GRE VSS reports means and standard deviations for all predictors and criteria. Scattergrams are presented for each predictor - criterion pair that has complete data for at least 25 students. Separate means and standard deviations are reported for the complete data pairs. The means based on pairs are compared with the means based on all students with data on each variable. Appendix A presents a sample of the presentation of each type of statistical information.

Analyses were replicated for subgroups based on sex, ethnic group, primary language, citizenship, and year of entry into the graduate program (up to nine entering classes could be combined to achieve the minimum sample size of 25), as long as there were at least 25 students in each analysis, and the subgroup consists of less than 95% of the total group.

WHAT ESTIMATION METHODS DO YOU USE?

The current GRE VSS uses ordinary least squares regression to produce correlation coefficients. In addition, if the removal of a small number of outliers changes the correlation by more than .2 or 1.96 standard errors (whichever is smaller), this is noted. Only the correlations based on the complete available data are entered into the distributions of correlations, however.

HOW DO YOU DEAL WITH STATISTICAL PROBLEMS PRESENTED BY SUBGROUPS OF STUDENTS, SMALL SAMPLES, HIGHLY INTERCORRELATED PREDICTORS, AND RESTRICTION OF RANGE?

Little is done systematically with regard to subgroups of students because few departments have enough data to perform any subgroup analyses. Fairly often, the subgroup of non-native English speakers poses problems. This group is particularly prevalent in departments in the physical and mathematical sciences. Although many of these students do very well in their graduate course work, a large number of them perform poorly on the GRE verbal measure. This may lead to low or negative validity coefficients. When this does occur it is discussed in the validity report. The Graduate Record Examinations are designed for examinees who are fluent in English.

The problem of small samples is handled in several ways. First, departments can combine up to nine years of data, as long as their coursework and grading standards have remained consistent. Second, the emphasis is on the interpretation of validity coefficients in the context of studies from other similar departments. Third, only simple descriptive statistics and zero-order correlations are reported. These statistics are relatively stable in the face of small sample sizes.

Highly intercorrelated predictors pose little problem for the current GRE VSS since statistical techniques that simultaneously consider several predictors, such as multiple regression, are not used.

Restriction of range and its general effect on validity coefficients is discussed in each validity report. Corrected validity coefficients are not reported for a number of reasons. First, we only have scores for the enrolled students. Typically data are no longer available for applicants when the departments decide to participate in the validity study service. Second, standard corrections for restriction of range do not appear appropriate for graduate school validity studies. Compensatory admissions policies (applicants with low GRE scores and low undergraduate grades do not bother applying to this graduate school, few applicants have high GRE scores and high UGPA and many of them go elsewhere even if accepted, we'll accept an applicant with low UGPA if he or she has high GRE scores or vice versa) lower the correlations between predictors in the group of enrolled students (Dawes, 1975). These correlations often become essentially zero or negative. This in turn lowers the correlations between predictors and criteria. Such an effect has been found in GRE VSS data (Kingston & Livingston, 1981) and hides the effect of standard restriction in range based on predictor variances (Gulliksen, 1950). See also Humphreys (1968) for additional concerns with the use of standard corrections for restriction of range.

WHAT SUMMARY INFORMATION (ACROSS INSTITUTIONS) DO YOU PROVIDE? HOW DO YOU ADVISE INDIVIDUAL USERS TO USE THE SUMMARY INFORMATION?

The GRE VSS has produced several summary reports (Kingston & Livingston, 1981; Livingston & Turner, 1982; Burton & Turner, 1983). Summary reports differ from the departmental reports primarily in that they provide correlations based on multiple predictors using graduate first-year average as a criterion, and summary data (primarily weighted mean correlations) for criteria other than first-year grade point average, most notably, faculty rating. Correlations based on multiple predictors have heretofore been produced two ways. A correlation matrix (V, Q, UGPA, FYA) was formed by taking the weighted mean correlations for each pair of variables. Multiple regression was performed on the resulting matrix, yielding beta weights and multiple correlations. In addition, somewhat arbitrarily weighted composites (unit weights for GRE V and Q and a weight of 300 for UGPA) were created and correlated with first-year grades. This was done for all participating departments combined and separately for all biological sciences departments, humanities departments, physical and mathematical sciences departments, and social science departments.

The summary reports have also provided evidence of the impact of compensatory admission policies on observed validity coefficients and the relationship between sample size and the variability of observed validity coefficients.

Individual users are not advised how to use the information provided in the summary report any more than they were explicitly told how to use the information in their individual reports. Basically, they are told whether or not a given predictor is related to performance in graduate school. For example, following are a few of the conclusions from the 1981-82 summary report (Livingston & Turner, 1982):

3. GRE scores tend to have moderate correlations with first-year grade averages. The best predictors are the GRE Advanced Test scores.

4. There are systematic differences between academic areas in the effectiveness of the various GRE scores for predicting first-year grades. The GRE quantitative score tends to be a much better predictor than the GRE verbal score in the mathematical and physical sciences, but not in the other academic areas.

6. GRE scores tend to predict faculty ratings of students more strongly than they predict first-year grade average.

THE GRE VSS TOMORROW

WHO USES THE TEST SCORES, FOR WHAT PURPOSE? WHO WILL USE THE VALIDITY INFORMATION, FOR WHAT PURPOSE?

This will not change under the revised GRE Validity Study Service.

WHAT TYPES OF STATISTICAL INFORMATION DO YOU REPORT? IN WHAT FORM DO YOU REPORT THEM?

The new GRE VSS will report three major types of statistical information: probability graphs, prediction equations, and expectancy tables. In addition, descriptive statistics for all measures will be reported using box-and-whisker graphs. All participants will be asked to pick two graduate first-year grade point averages. For the first, they are to decide what is the lowest GPA that could be obtained by any student that they would still want to have in the program. Perhaps such a student would be creative but would do poorly on classroom tests or would be a source of cultural or philosophical diversity, and thus would still be a desirable addition to the program even with relatively low grades. The second GPA selected by the department would be one beyond which other factors would be far more important than any additional increments in grades. Similarly, for other criteria that a department would use in their study, two scores representing minimally acceptable and clearly acceptable performance would be selected.

These two GPAs would be used to produce probability graphs. For each predictor -- including but not limited to GRE verbal (V), quantitative (Q), analytical (A), GRE Subject Tests (S), and undergraduate grades (U) -- and for composites of V, Q, A, and U and V, Q, A, U, and S, graphs would be produced. The horizontal axis of each graph will be the score on the measure or composite. The vertical axis will be the probability of exceeding the chosen GPA. Examples of these graphs can be found in Appendix B.

Prediction equations, expressed with two significant digits, will be produced for each predictor separately and for the two composites. In addition, a table of expected criterion scores for each of a number of predictor and composite values will be produced. Examples of such tables can also be found in Appendix B.

Also, box-and-whisker plots will be presented for each predictor and criterion. Once again, see Appendix B for examples.

WHAT ESTIMATION METHODS DO YOU USE?

Empirical Bayes regression will be used to determine the relationships between multiple predictors and criteria. The particular empirical Bayes method to be used by the GRE VSS is a multistep procedure. First, ordinary least squares methods are used to estimate the regression of the criterion on the predictors. This is done separately for each participating department. Collateral information, in particular the mean score on each predictor, is also gathered for each department. Then another least squares regression is performed; the departmental regression weights are themselves regressed on the collateral information. This new equation allows us to predict regression weights from departmental characteristics. Then, the least squares regression weights are shrunken toward the predicted regression weights. How far they are shrunken depends on the standard error of the least squares regression weights. The less sure we are of the least squares regression weights, perhaps because they are based on a very small sample or very collinear predictors, the more they are shrunken towards the predicted regression weights. Braun and Jones (1985) have shown this procedure to work well. Regression weights show less year to year variation within the same department, are more similar to each other than least squares weights are in similar departments (for example, across literature departments), and most important, perform better upon cross-validation.

Multiple correlations will be estimated as the square root of the quotient of the sum of the squared deviations from the Bayesian fit (that is, observed criterion score minus predicted criterion score) and the sum of the squared deviations from the mean (that is, observed criterion score minus mean criterion score). This is exactly the same as with a least squares correlation, except that the predicted criterion score is based on the empirical Bayes regression coefficients.

HOW DO YOU DEAL WITH STATISTICAL PROBLEMS PRESENTED BY SUBGROUPS OF STUDENTS, SMALL SAMPLES, HIGHLY INTERCORRELATED PREDICTORS, AND RESTRICTION OF RANGE?

Subgroups of students will be handled in two ways. The only subgroups for which separate analyses will be performed at the departmental level will be based on language fluency. In recognition of the fact that the GRE tests are designed for fluent speakers of English, the new GRE VSS will not combine data for students whose best language is English with those whose best language is not English. Separate studies may be done for these two groups if departments so desire, as long as they have sufficient students in each group for which a study is performed. "Sufficient students" has not yet been defined, but it is expected to be somewhere between 5 and 10.

It is not expected that there will be sufficient data to estimate separate regression equations for other subgroups, even using empirical Bayes methods. Instead, occasional summary reports will compare residuals for

various subgroups of interest such as Blacks, Whites, males, females, older students, or science undergraduate majors in social science graduate fields.

The problem of small samples will be handled by using empirical Bayes regression to determine the relationships between the predictors and criteria.

As of yet there are no plans to explicitly account for restriction of range in the new GRE VSS.

WHAT SUMMARY INFORMATION (ACROSS INSTITUTIONS) DO YOU PROVIDE? HOW DO YOU ADVISE INDIVIDUAL USERS TO USE THE SUMMARY INFORMATION?

The development of summary reports based on empirical Bayes methods is still in the planning stage. In addition to mean or median correlations and multiple correlations based on subgroups of similar departments (such as, anthropology, biology, computer science, English, etc.) and analysis of residuals for members of various subgroups of examinees, it is not clear what analyses might be done.

REFERENCES

Braun, H. and Jones, D. (1985). The use of empirical Bayes methods in the study of the validity of academic predictors of graduate school performance. GRE Board Professional Report No. 79-13P. Princeton, NJ: Educational Testing Service.

Burton, N. W. and Turner, N. J. (1983). Effectiveness of the Graduate Record Examinations for predicting first-year grades: 1981-82 summary report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Dawes, R. M. (1975). Graduate admission variables and future success. Science, 187, 721-723.

Educational Testing Service (1986). Guide to the use of the Graduate Record Examinations, 1986-87. Princeton, NJ: Educational Testing Service.

Gulliksen, H. (1950). Theory of mental tests. New York: Wiley.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. Journal of Educational Psychology, 59, 375-380.

Kingston, N. M. and Livingston, S. (1981). Effectiveness of the Graduate Record Examinations for predicting first-year grades: 1979-80 summary report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Livingston, S. A. and Turner, N. J. (1982). Effectiveness of the Graduate Record Examinations for predicting first-year grades: 1980-81 summary report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1979). The validation of GRE scores as predictors of first-year performance in graduate study: Report of the GRE cooperative validity study project. GRE Board Research Report No. 75-8R. Princeton, NJ: Educational Testing Service.

APPENDIX A

TABLE 4

SUMMARY STATISTICS FOR INDIVIDUAL VARIABLES

VARIABLE	MINIMUM OBSERVED	MAXIMUM OBSERVED	MEAN	STANDARD DEVIATION	NUMBER OF STUDENTS
GRE VERBAL.....	220	710	493.8	109.5	26
GRE QUANTITATIVE.....	540	770	628.1	61.1	26
UNDERGRADUATE GPA.....	1.33	3.84	3.283	0.503	25
GRADUATE FIRST-YEAR GPA.....	1.64	3.88	3.467	0.494	26

TABLE 5

PREDICTOR-CRITERION CORRELATIONS
(AND NUMBERS OF STUDENTS ON WHICH THEY ARE BASED)

PREDICTOR	FIRST-YEAR GRADUATE GPA
GRE VERBAL.....	0.31 (26)
GRE QUANTITATIVE.....	-0.22 (26)
UNDERGRADUATE GPA.....	0.61 (25)

02/24/86

GRADUATE RECORD EXAMINATIONS
VALIDITY STUDY SERVICE

DISTRIBUTION OF CORRELATIONS OF GRE VERBAL ABILITY SCORE
WITH FIRST-YEAR GRADE-POINT AVERAGE FOR
BIOLOGICAL SCIENCES DEPARTMENTS

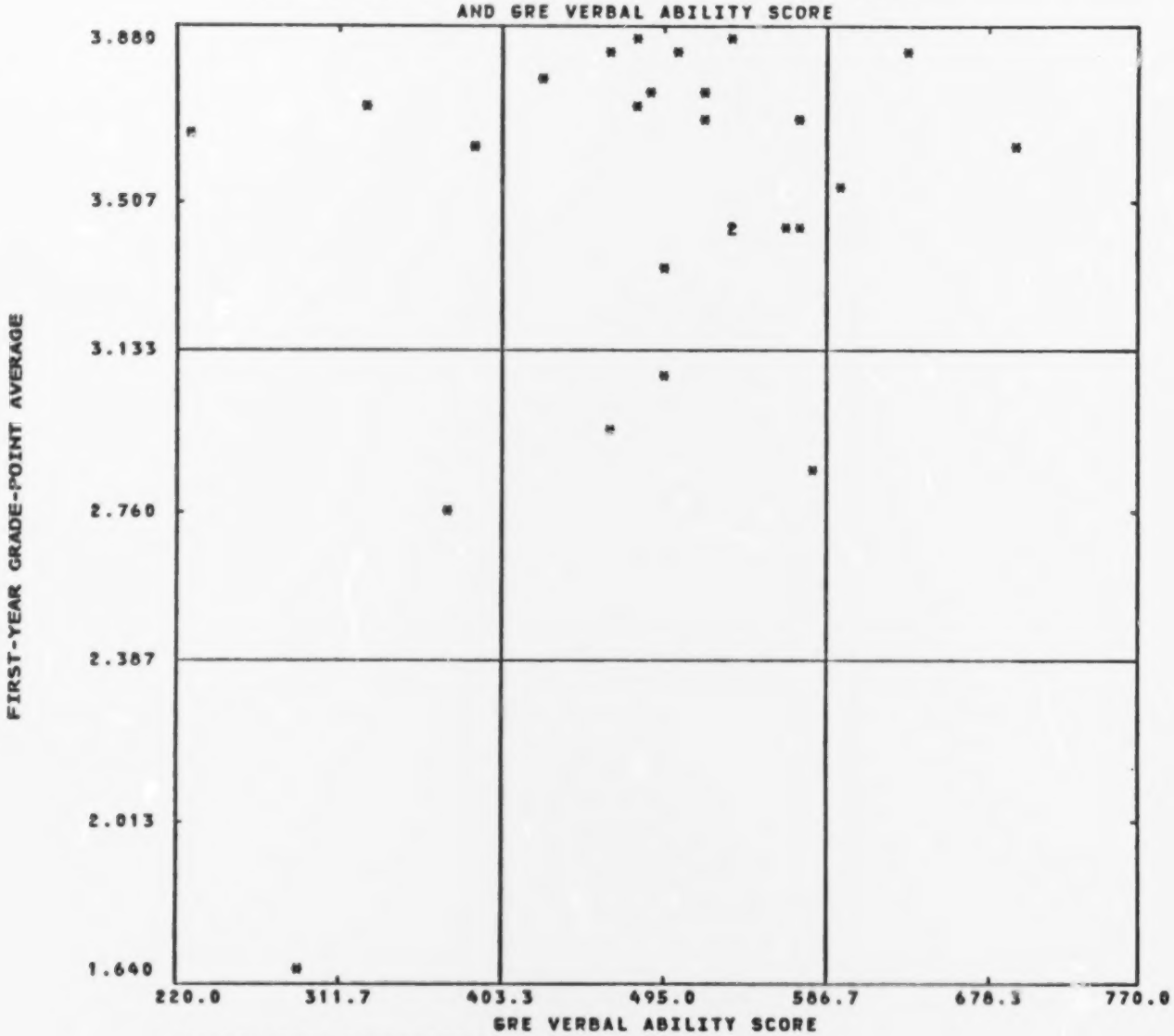
CORRELATION	FREQUENCY
.90 TO .99	
.80 TO .89	
.70 TO .79	
.60 TO .69	::
.50 TO .59	:::
.40 TO .49	::
.30 TO .39	X:::::::::::.
.20 TO .29	:::::::::::::.
.10 TO .19	::::::::::.
.00 TO .09	:::::
-.10 TO -.01	::::
-.20 TO -.11	::
-.30 TO -.21	.
-.40 TO -.31	.
-.50 TO -.41	
-.60 TO -.51	
-.70 TO -.61	
-.80 TO -.71	
-.90 TO -.81	
-.99 TO -.91	

A DOT REPRESENTS ONE DEPARTMENT.

THE X LOCATES YOUR DEPARTMENT'S CORRELATION COEFFICIENT.

DATA ARE BASED ON 107 VALIDITY STUDIES CONDUCTED BETWEEN 1976 AND 1986.

FIGURE 10
SCATTER DIAGRAM OF FIRST-YEAR GRADE-POINT AVERAGE
AND GRE VERBAL ABILITY SCORE



CORRELATION COEFFICIENT = 0.31

STUDENTS HAVING DATA ON BOTH VARIABLES	MEAN	S.D.	N
FIRST-YEAR GRADE-POINT AVERAGE	3.467	0.494	26
GRE VERBAL ABILITY SCORE	493.8	109.5	

STUDENTS HAVING DATA ON EITHER VARIABLE	MEAN	S.D.	N
FIRST-YEAR GRADE-POINT AVERAGE	3.467	0.494	26
GRE VERBAL ABILITY SCORE	493.8	109.5	26

APPENDIX B

Part B: Relationships of GRE Scores and Undergraduate Grade-Point Average to Students' Typical FYA.

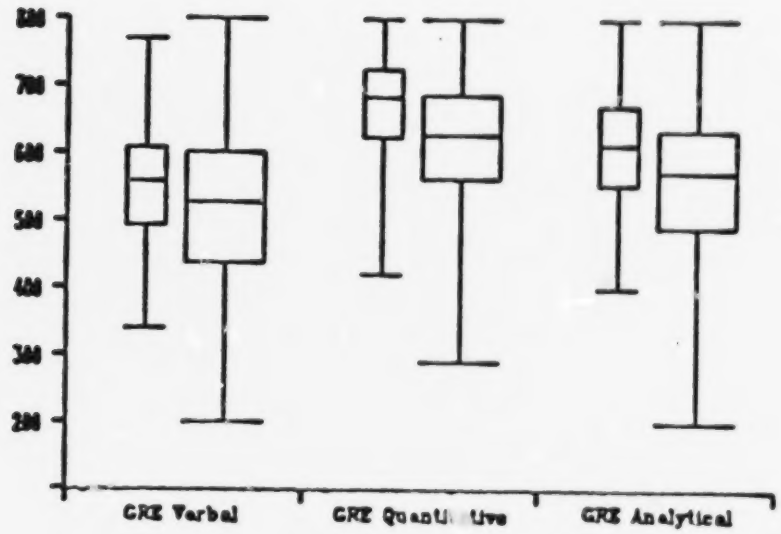
The following table shows estimates of the average FYA for a large group of applicants typical of the students in the Department of Physics at Anytown University, with a given GRE score, undergraduate grade-point average, or a given value on the Composite Score.

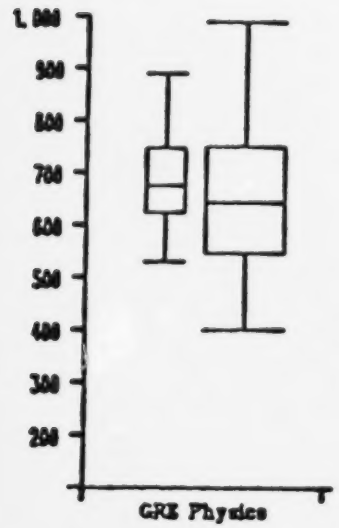
Predicted Average FYA
for the Department of Physics at Anytown University

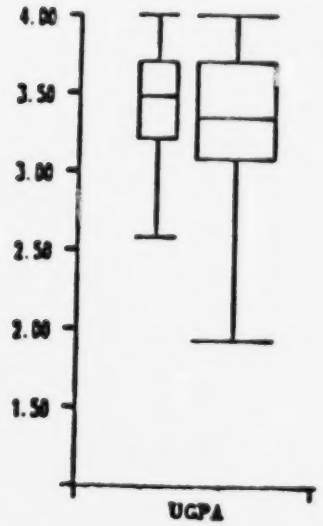
GRE V	Avg. FYA	GRE Q	Avg. FYA	GRE A	Avg. FYA	GRE S	Avg. FYA	UGPA	Avg. FYA	Composite Score[1]	Avg. FYA
						950	3.9				
						900	3.8				
						850	3.7				
800	3.7	800	3.7	800	3.6	800	3.7	4.0	3.6	2000	4.0
750	3.6	750	3.6	750	3.6	750	3.6	3.8	3.6	1900	3.8
700	3.6	700	3.5	700	3.5	700	3.5	3.6	3.5	1800	3.6
650	3.5	650	3.4	650	3.5	650	3.4	3.4	3.5	1700	3.4
600	3.5	600	3.3	600	3.4	600	3.3	3.2	3.4	1600	3.2
550	3.5	550	3.2	550	3.4	550	3.2	3.0	3.4	1500	3.0
500	3.4	500	3.2	500	3.3	500	3.1	2.8	3.3	1400	2.8
450	3.4	450	3.1	450	3.3	450	3.1	2.6	3.3	1300	2.6
400	3.3	400	3.0	400	3.2	400	3.0	2.4	3.2	1200	2.4
350	3.3	350	2.9	350	3.2	350	2.9	2.2	3.1	1100	2.2
300	3.2	300	2.8	300	3.1	300	2.8	2.0	3.1	1000	2.0
250	3.2	250	2.7	250	3.1	250	2.7	1.8	3.0		
200	3.1	200	2.6	200	3.0	200	2.6	1.6	3.0		

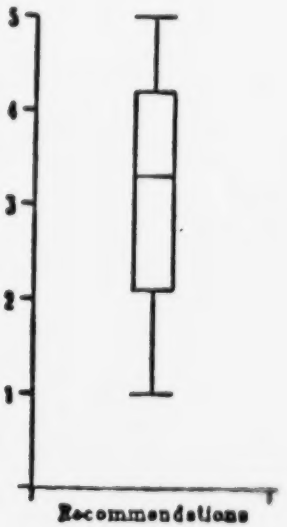
[1] Composite scores tend to give more accurate estimates of FYA when based on more admission measures, but the average FYA is the same for a given composite score regardless of which formula was used to develop it.

Distributions of Admission Measure Scores

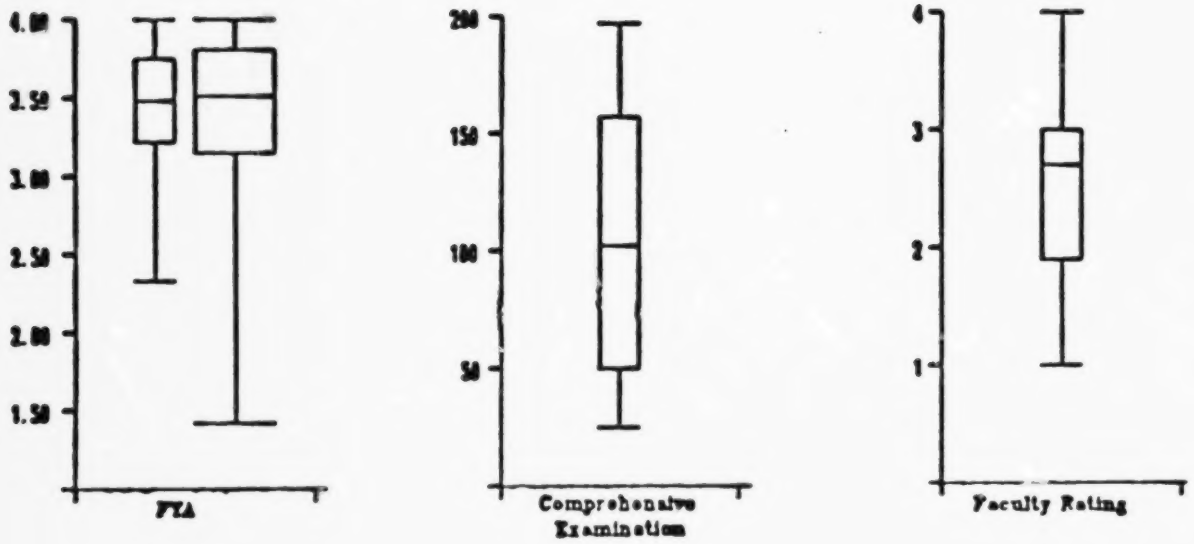








Distributions of Criterion Measure Scores



Probability of Earning at Least a 3.0 or 3.5 FYA Based on Applicant GRE Verbal, Quantitative, and Analytical Scores